<Bioinformatics Center>Chemical Lfe Science

Bioinformatics Center
– Chemical Life Science –

http://cls.kuicr.kyoto-u.ac.jp/

Assoc Prof
GOTO, Susumu
(D Eng)

Program-Specific Assist Prof
TOKIMATSU, Toshiaki
(D Agr)

Program-Specific Assist Prof
KOTERA, Masaaki
(D Sc)

PD
KIRWAN, Gemma May
(Ph D)

PD
JOANNIN, Nicolas
(Ph D)

Researchers

MORIYA, Yuki
NAKAGAWA, Zenichi
MUTO, Ai

Students

SHIMIZU, Yugo (D3)
MIZUTANI, Sayaka (D3)
HIRANUKA, Kazushi (D3)
NISHIMURA, Yosuke (D3)
FUJITA, Megumi (D3)
MIHARA, Tomoko (D2)
JIN, Zhao (D2)
CHEN, Zhang (M2)
SATOH, Takanori (M1)
NORO, Yousuke (UG)

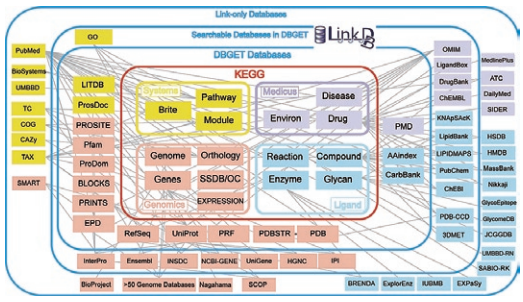
Visitor

Dr TSAI, Yu-Shuen National Yang-Ming University, Taiwan, 7 October 2011–30 April 2012

Scope of Research

The proteins responsible for biosynthesis, biodegradation, and transport of additional molecules, such as small metabolites, lipids and glycans, are encoded in the genome, which may indicate that all cellular functions are specified by the genomic DNA sequence. In practice, however, inferring higher-level systemic functions of the cell or the organism needs more than solely the genomic information. We are developing bioinformatics methods to integrate different types of data and knowledge on various aspects of the biological systems towards basic understanding of life as a molecular interaction/reaction system and also toward practical applications in medical and pharmaceutical sciences.

GenomeNet Top page



Databases available in the DBGET/LinkDB system of the GenomeNet service. Color of each database represents the type of its contents, yellow: systems information, purple: medical information, pink: genetic information, light blue: chemical information.

KEYWORDS

GenomeNet
(Meta)genomics
Pathway
Bioinformatics
Pharmacoinformatics

Selected Publications

Takarabe, M.; Kotera, M.; Nishimura, Y.; Goto, S.; Yamanishi, Y., Drug Target Prediction Using Adversed Event Report Systems: A Pharmacogenomic Approach, *Bioinformatics*, **28,** i611-i618 (2012).

Mizutani, S.; Pauwels, E.; Stoven, V.; Goto, S.; Yamanishi, Y., Relating Drug-protein Interaction Network with Drug Side-effects, *Bioinformatics*, **28,** i522-i528 (2012).

Kotera, M.; Yamanishi, Y.; Moriya, Y.; Kanehisa, M.; Goto, S., GENIES: Gene Network Inference Engine Based on Supervised Analysis, *Nucleic Acids Res.*, **40,** W162-W167 (2012).

Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M., KEGG for Integration and Interpretation of Large-scale Molecular Datasets, *Nucleic Acids Res.*, **40,** D109-D114 (2012).

Takarabe, M.; Shigemizu, D.; Kotera, M.; Goto, S.; Kanehisa, M., Network-based Analysis and Characterization of Adverse Drug-drug Interactions, *J. Chem. Inf. Model.*, **51,** 2977-2985 (2011).

Development of a Web-based Supervised Gene Network Inference Engine

Recent developments of biotechnologies, e.g., microarray and proteomics technologies, contribute to an increasing amount of high-throughput data for genes and proteins, which are useful sources to infer the biological networks on a large scale. We have developed GEne Network Inference Engine based on Supervised analysis (GENIES: http://www.genome.jp/tools/genies/), a web server to predict unknown part of gene network from various types of genome-wide data in the framework of supervised network inference (Kotera et al., 2012). The originality of the supervised network inference method lies in construction of a predictive model using partially known network information, and in the integration of heterogeneous data with kernel methods. The method is suitable for predicting potential interactions involving uncharacterized genes and their associations with known pathways. The GENIES server accepts any profiles of genes or proteins (such as gene expression profiles, protein subcellular localization profiles and phylogenetic profiles), or pre-calculated gene-gene similarity matrices (or kernels) in the tab-delimited file format. As a training data set to learn a predictive model, the users can choose either known molecular network information in the KEGG Pathway database, or their own gene network data. The server provides the list of newly predicted gene pairs, maps the predicted gene pairs onto the associated pathway diagrams in KEGG Pathway, and indicates candidate genes for missing enzymes in organism-specific metabolic pathways.

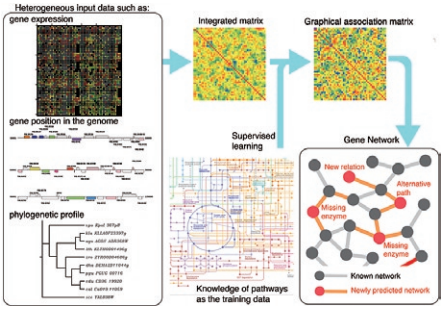


Figure 1. The workflow of GENIES.

A Statistical Framework for Analyzing the Association between Drug Target Proteins and Drug Side-effects

Most drugs are small chemical compounds that target specific proteins to induce regulations of the overall protein networks, or pathways. This may result in expected therapeutic effects to the body, or occasionally undesired side-effects. Recently, several types of drug-related data have been accumulated in publicly available databases. Drug chemical structures, target proteins, and side-effects are such data types. Bioinformatics approaches are expected to be useful to retrieve important relationships between different types of data, which helps studies on the underlying mechanisms of drug-actions in a living system at different levels. Our lab has been focusing on storing such heterogeneous data in the form of relational database and developing related algorithms. Previously, we published a drug-drug interaction retrieval system in the KEGG DRUG database, which is a database-searching approach to comprehensively understand drug-drug interactions at the molecular scale (Takarabe et al., 2011). We took another approaches applying statistical methods, which link drug-protein interaction data at the molecular scale and their side-effects at the phenotypic scale. As one of our recent studies (Mizutani et al., 2012), we applied the sparse canonical correlation analysis (SCCA) to a drug-protein interaction data in a framework of predicting potential side-effects for a set of drugs (Figure 2). This resulted in an extraction of target proteins and side-effects in the form of correlated sets (Figure 3). A following pathway enrichment analysis using KEGG and Gene Ontology (GO) databases showed that proteins of similar biological processes were clustered together in the same extracted correlated sets, even if their molecular functions were different. In another study (Takarabe et al., 2012), we used similar approach to retrieve biologically relevant interpretations from the drug-target-side-effect associations. We believe our continuous efforts help promote discussions over the molecular mechanisms of drug side-effects and underlying biological knowledge.

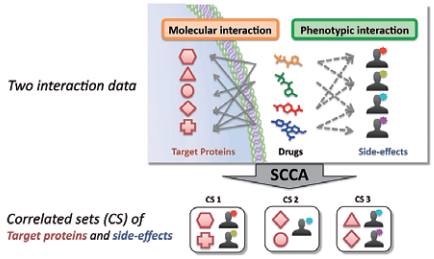


Figure 2. SCCA extracts correlated sets of target proteins and side-effects.

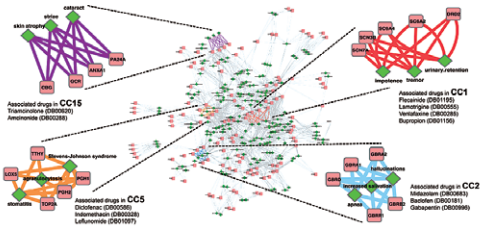


Figure 3. An illustration of the network of drug targeted proteins and side-effects in the extracted correlated sets (CSs).